Open Access Full Text Article

ORIGINAL RESEARCH

Development and validation of risk score to estimate 1-year late poststroke epilepsy risk in ischemic stroke patients

Nai-Fang Chi[1,2]
Yi-Chun Kuan[1–3]
Yao-Hsien Huang[1,2]
Lung Chan[1,2]
Chaur-Jong Hu[1,2,4]
Hung-Yi Liu[5]
Hung-Yi Chiou[6]
Li-Nien Chien[7]

[1]Department of Neurology, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan, Republic of China; [2]Department of Neurology, Shuang Ho Hospital, Taipei Medical University, Taipei, Taiwan, Republic of China; [3]Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan, Republic of China; [4]PhD Program for Neural Regenerative Medicine, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan, Republic of China; [5]Health and Clinical Research Data Center, College of Public Health, Taipei Medical University, Taipei, Taiwan, Republic of China; [6]School of Public Health, College of Public Health, Taipei Medical University, Taipei, Taiwan, Republic of China; [7]School of Health Care Administration, College of Management, Taipei Medical University, Taipei, Taiwan, Republic of China

Correspondence: Li-Nien Chien
School of Health Care Administration, College of Management, Taipei Medical University, No 172-1, Section 2, Keelung Road, Da'an District, Taipei City 106, Taiwan, Republic of China
Tel +886 2 6638 2736
Fax +886 2 2378 9788
Email lnchien@tmu.edu.tw

Objective: This study aimed to develop and validate a prognostic model for the 1-year risk of late poststroke epilepsy (PSE).

Materials and methods: We included patients initially diagnosed with ischemic stroke between 2003 and 2014 in a National Health Insurance claims-based cohort in Taiwan. Patients were further divided into development and validation cohorts based on their year of stroke diagnosis. Multivariable Cox regression with backward elimination was used to analyze the association between 1-year PSE and risk factors before and on stroke admission.

Results: In total, 1,684 (1.93%) and 725 (1.87%) ischemic stroke patients comprising the development and validation cohorts, respectively, experienced late PSE within 1 year after stroke. Seven clinical variables were examined to be independently associated with 1-year risk of PSE. We developed a risk score called "PSEiCARe" ranging from 0 to 16 points, comprising the following factors: prolonged hospital stay (>2 weeks, 1 point), seizure on admission (6 points), elderly patients (age ≥80 years, 1 point), intensive care unit stay on admission (3 points), cognitive impairment (dementia, 2 points), atrial fibrillation (2 points), and respiratory tract infection (pneumonia) on admission (1 point). Patients were further classified into low-, medium-, high-, and very-high-risk groups. The incidence (per 100 person-years) was 0.64 (95% CI: 0.56–0.71) for the low-risk, 2.62 (95% CI: 2.43–2.82) for the medium-risk, 10.3 (95% CI: 9.48–11.3) for the high-risk, and 28.2 (95% CI: 24.0–33.0) for the very-high-risk groups. Discrimination and calibration were satisfactory, with a Harrell's C of 0.762 in the development model and 0.792 in the validation model.

Conclusion: PSEiCARe is an easy-to-use prognostic score that integrates patient characteristics and clinical factors on stroke admission to predict 1-year PSE risk; it has the potential to assist individualized patient management and improve clinical practice, thereby preventing the occurrence of late PSE.

Keywords: poststroke epilepsy, prediction model, risk score, population-based, claim analysis

Introduction

Stroke is a major cause of epilepsy in adults, and elderly people in particular.[1] After stroke, seizures and subsequent development of epilepsy may impede patient recovery because rehabilitation is delayed and the risk of complications such as fall-related injuries is higher.[2] Moreover, seizure frequency directly impairs a patient's quality of life.[3] Health care systems are facing an increasing number of elderly people with poststroke seizures because of demographic changes worldwide.[4] Therefore, determining and managing the risk factors of poststroke epilepsy (PSE) is crucial for stroke care.

Several factors, such as hemorrhagic stroke, cortical lesion, stroke severity, and early seizures, are risk factors for late-onset PSE.[5–8] Poststroke infection and atrial fibrillation are also associated with late PSE.[2,9,10] In patients with high risk factors for late PSE, such as aneurysmal subarachnoid hemorrhage, prophylactic antiepileptic drugs (AEDs) are usually prescribed.[11,12] In patients with ischemic stroke, AED use is only recommended after clinical seizure onset.[13] The benefit of prophylactic AED use in both hemorrhagic and ischemic stroke remains unclear because randomized control trials of late PSE prophylaxis and studies of late PSE risk estimation are lacking. Therefore, a late PSE risk-predicting score would be helpful in determining prevention strategies such as the use of prophylactic AEDs. However, precisely estimating the risk of late PSE is challenging because the frequency of late PSE is 2%–4%.[7] Such predictions are especially difficult for patients with ischemic stroke because the incidence of late PSE in ischemic stroke is lower than that in hemorrhagic stroke.[14] Thus far, only 1 study, which enrolled 246 patients with stroke, has proposed and developed a late PSE risk scale; it uses type, location, and severity of stroke, as well as the onset of poststroke seizure, to predict the risk of PSE within 1 year of a stroke.[6] The performance of this scale has not been validated, and it is not generally used in clinical practice.[15,16]

Improving risk prediction accuracy and exploring novel risk factors in a larger population may enable more targeted prophylaxis. Furthermore, developing a simple risk score that integrates data obtained in the acute stage of stroke is imperative for quantifying the late PSE risk. Therefore, this study developed and validated a scoring system for predicting 1-year late PSE risk using a population-based health care claims database of ischemic stroke patients.

Materials and methods
Ethics statement
This study was approved by the Taipei Medical University Joint Institutional Review Board (TMU-JIRB, approval number 201502017). Patients were selected from the National Health Insurance Research Database (NHIRD) of Taiwan, which is maintained by the Health and Welfare Data Science Center (HWDC), Ministry of Health and Welfare, Executive Yuan, Taiwan. To protect patients' privacy, the HWDC encrypts individual identifiers before releasing information to researchers. Informed consent of the patients was thus waived in the full TMU-JIRB review process.

Data sources
In this population-based cohort study, we extracted data from the NHIRD for the period 2000–2015. The NHIRD is managed by the National Health Insurance Administration and covers almost all medical service claims received from beneficiaries. Under law, every resident of Taiwan must enroll in the national health insurance; the rate of coverage was 99% in 2016. Death records were obtained from the National Death Registry, a population-based registry of cause of death in Taiwan. Death records in Taiwan are highly comprehensive and accurate because it is mandatory under law to register all deaths. Furthermore, all residents of Taiwan have a permanent personal identification number that links many national databases at an individual level. The 2 cohorts in this study were linked using unique encrypted identifiers regulated by the HWDC and Personal Information Protection Act.

Participants
We included all patients diagnosed with ischemic stroke from 2003 to 2014. These patients were defined on the basis of a discharge diagnosis of ischemic stroke (International Classification of Diseases, Ninth Revision, Clinical Modification [ICD-9-CM]: 433.xx, 434.xx, and 436.xx). A review period of at least 3 years was adopted to exclude patients with a history of stroke, epilepsy, and the following known risks of epilepsy other than stroke: traumatic brain injury, metastatic carcinoma, brain tumor, anoxic brain injury, encephalopathy, alcohol abuse, drug abuse, and central nervous system infections. Thus, patients in whom the cause of epilepsy was primarily due to stroke were included. Detailed information on disease diagnostic coding is provided in Table S1. We excluded patients younger than 45 years because stroke etiologies in young patients differ from those in middle-aged and elderly patients.[17]

The final study cohort comprised 125,757 patients with ischemic stroke. Because the sample size was large, we divided the patients according to their year of stroke diagnosis. The development cohort (N=87,068) comprised patients registered within the first 8 years; the validation cohort (N=38,689) comprised the remaining patients. The basic characteristics of the 2 cohorts are listed in Table 1.

Outcomes
In Taiwan, diagnosis of seizures usually follows the recommendations of the International League Against Epilepsy (ILAE). A seizure occurring within 7 days of a stroke is defined as an "early seizure", whereas ≥2 seizures occurring

Table 1 Patient characteristics used to develop and validate the risk scoring system

Characteristics	Overall		Development cohort		Validation cohort		P
	N	(%)	N	(%)	N	(%)	
Sample size	125,757	–	87,068		38,689		–
Male, yes (%)	74,585	(59.3)	51,026	(58.6)	23,559	(60.9)	<0.001
Age in years, mean (SD)	68.6	(11.7)	68.7	(11.6)	68.5	(11.8)	0.004
Age group, years							<0.001
45–54	17,283	(13.7)	12,016	(13.8)	5,267	(13.6)	–
55–64	29,981	(23.8)	19,613	(22.5)	10,368	(26.8)	–
65–79	53,890	(42.9)	38,733	(44.5)	15,157	(39.2)	–
≥80	24,603	(19.6)	16,706	(19.2)	7,897	(20.4)	–
Stroke conditions at admission							
LOS in days, mean (SD)	9.7	(7.8)	9.7	(7.8)	9.8	(7.7)	0.001
LOS >2 weeks, yes (%)	25,119	(20.0)	17,193	(19.7)	7,926	(20.5)	0.003
ICU stay in days, mean (SD)	1.27	(3.83)	1.27	(3.95)	1.28	(3.57)	0.559
ICU stay, yes (%)	23,579	(18.7)	15,331	(17.6)	8,248	(21.3)	<0.001
Seizure, yes (%)	6,547	(5.2)	4,826	(5.5)	1,721	(4.4)	<0.001
Pneumonia, yes (%)	11,258	(9.0)	7,427	(8.5)	3,831	(9.9)	<0.001
Preexisting disease conditions, yes (%)							
Hypertension	93,301	(74.2)	64,216	(73.8)	29,085	(75.2)	<0.001
Diabetes	48,485	(38.6)	33,088	(38.0)	15,397	(39.8)	<0.001
Hyperlipidemia	45,597	(36.3)	27,813	(31.9)	17,784	(46.0)	<0.001
Coronary artery disease	20,522	(16.3)	14,983	(17.2)	5,539	(14.3)	<0.001
Heart failure	10,464	(8.3)	7,338	(8.4)	3,126	(8.1)	0.039
Atrial fibrillation	15,266	(12.1)	9,886	(11.4)	5,380	(13.9)	<0.001
Peripheral artery disease	1,118	(0.9)	755	(0.9)	363	(0.9)	0.215
Malignant neoplasm	7,595	(6.0)	4,952	(5.7)	2,643	(6.8)	<0.001
Renal disease	8,496	(6.8)	5,724	(6.6)	2,772	(7.2)	0.001
Rheumatoid arthritis or collagen vascular disease	2,498	(2.0)	1,649	(1.9)	849	(2.2)	<0.001
Dementia	3,563	(2.8)	2,405	(2.8)	1,158	(3.0)	0.023

Note: P-values were calculated based on a Student's t-test for continuous variables and chi-squared test for categorical variables.
Abbreviations: LOS, length of hospital stay; ICU, intensive care unit.

>24 hours apart later than 7 days or more poststroke are defined as "late PSE".[18] The primary end point in this study was the occurrence of late PSE within 1 year of follow-up. Because the exact time of seizures was not recorded in the NHIRD, we used the following definitions of early seizure and late PSE to fulfill the ILAE criteria as far as possible. An early seizure occurred if the patient had a diagnosis of epilepsy or convulsion (ICD-9-CM: 345.x, 780.3, or 780.39) on stroke admission. Late PSE was assumed if patients had a diagnosis of epilepsy (ICD-9-CM: 345.x, which represented at least 2 seizures before the visit) or at least 2 claims of convulsions (ICD-9-CM: 780.3 or 780.39) after the stroke admission on an emergency department visit or hospital admission at different dates; or if patients had an AED prescription claim. The latter criterion was meant to ensure that the epilepsy was a new episode by limiting the study cohort to patients with no AED claims during at least the 3 months before late PSE development. Patients not diagnosed with epilepsy or who died within 1 year of follow-up were considered censored cases.

Predictors

Participant variables representing 1-year late PSE risk predictors included sex, age, early seizure on stroke admission, pneumonia occurring after stroke, and preexisting atrial fibrillation, which are known risk factors for late PSE.[2,5–7,9,10] Because stroke severity was not recorded in the NHIRD, we used intensive care unit (ICU) stay and length of hospital stay (LOS) of >2 weeks, which was the uppermost quartile for all patients with stroke, as a proxy for stroke severity on admission. Moreover, we considered other potential risk factors that may affect stroke severity or outcomes: comorbidities of hypertension, diabetes, hyperlipidemia, coronary heart disease, heart failure, peripheral arterial disease, malignant neoplasms, renal disease, moderate or severe liver disease, rheumatoid arthritis or collagen vascular disease, and dementia.[19]

Comorbidities were derived from patients' medical claims. The comorbidity observation period was 11 months before and 1 month after stroke admission. To increase the validity of diagnoses, patients were identified as having a

specific disease if they received the same diagnosis during 2 clinic or hospital visits within the observation period, with a 30-day interval between the 2 visits. Although cortical lesion is a risk factor for late PSE, we could not include it because the NHIRD does not contain information on stroke lesion location.

Statistical analyses

All covariates were entered into a multivariate Cox regression model as dichotomous variables to make the model easy for clinicians to use. We also evaluated each independent variable for adherence to the assumption of proportional hazards. Predictors were included in the final model on a backward elimination basis, with *P*-values of <0.05 and >0.1 determining inclusion and removal, respectively.

One goal of this study was to develop an easy-to-use prognostic score with which clinicians can stratify patient risk. Therefore, the patients were divided into risk groups by summing the independent predictors, which were weighted by their corresponding β-coefficients based on the scoring system of Schneeweiss,[20] to produce a risk score ranging from 0 to 16. A similar approach was adopted when rounding up independent predictors to the nearest integer, resulting in the 4 predictors of ICU stay, seizure on stroke admission, atrial fibrillation, and cognitive impairment (dementia), with the score ranging from 0 to 5.

Validation of the prediction model

The model performance was measured according to Harrell's *C* index of concordance, which is commonly used to evaluate a prediction model's discriminatory power based on the Cox proportional-hazards model.[21,22] We assessed internal validity with a bootstrapping validation approach to account for overfitting in the entire model development process. We repeated the entire model selection process in 100 samples drawn with replacement from the development sample.

Model fit was also assessed using the validation data set. The validation set was created using a temporal modeling approach that retrieved data from the same database but from a later study period. This approach was recommended by the transparent reporting of a multivariate prediction model for individual prognosis or diagnosis (TRIPOD) group as an intermediate between internal and external validation.[23]

Results
Participants

Of the patients with new onset of stroke but without a known risk of epilepsy, 1,684 (1.93%) and 725 (1.87%) in the development and validation cohorts, respectively, developed late PSE within 1 year of follow-up (Figure 1).

In the development cohort, there were more male (58.6%) than female patients, and the mean age of the patients with stroke was 68.6 years, with a standard deviation of 11.6 years. Regarding stroke admission status, 19.7% had a LOS of >2 weeks (upper quartile), 17.6% had an ICU stay, 5.5% had seizure and 8.5% pneumonia on stroke admission, and 11.4% had preexisting atrial fibrillation. Furthermore, regarding the late PSE risk factors, 73.8% of the patients had hypertension, 38.0% diabetes, 31.9% hyperlipidemia, 17.2% coronary artery disease, 8.4% heart failure, 6.6% renal failure, 5.7% malignant neoplasm, 2.8% dementia, 1.9% rheumatoid arthritis or collagen vascular disease, and 0.9% peripheral artery disease.

Compared with the development cohort, the validation cohort had a higher proportion of patients who were male, aged 55–64 years, and diagnosed with pneumonia on stroke admission and who had hypertension, diabetes, hyperlipidemia, atrial fibrillation, and malignant neoplasm, but with a lower proportion of those with coronary artery disease and heart failure (Table 1).
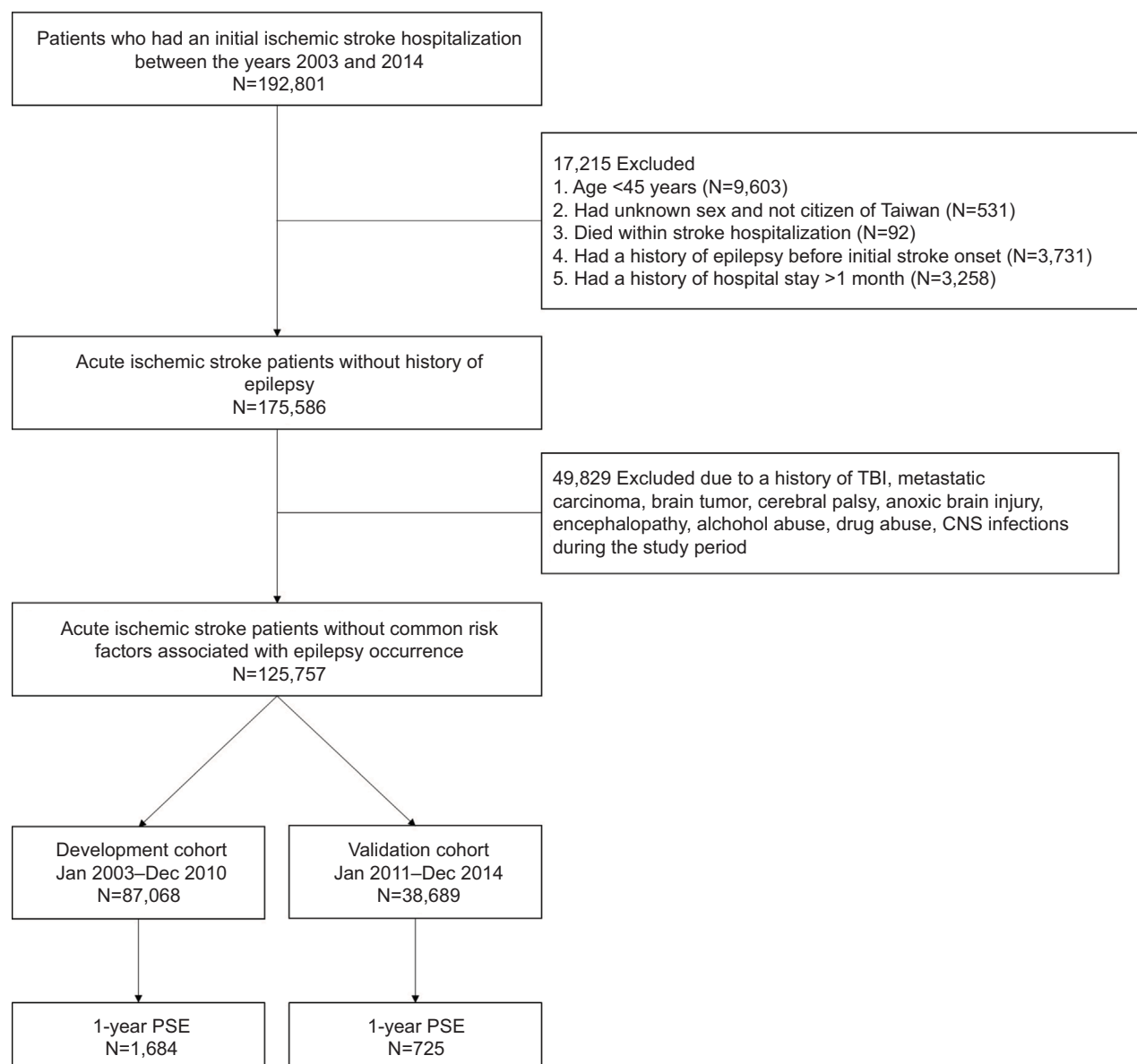
Model development

According to multivariate Cox regression analysis, the independent predictors associated with 1-year late PSE risk were age ≥80 years, LOS >2 weeks, ICU stay, early seizure, pneumonia on stroke admission, preexisting atrial fibrillation, and dementia (Table 2).

To make the predictive model easy for clinicians to use, we adopted a risk scoring system with the acronym "PSEi-CARe", consisting of "Prolonged hospital stay (>2 weeks)", "Seizure on stroke admission", "Elderly patients (age ≥80 years)", "ICU stay", "Cognitive impairment (dementia)", "Atrial fibrillation", and "Respiratory tract infection (pneumonia) on stroke admission". Two scoring systems were considered. The first method was based on scoring each predictor by rounding the corresponding β-coefficients in the regression model up to the nearest integer, thus creating a 5-point scoring system that is simple from a clinical practice perspective. The second method, developed by Schneeweiss, was relatively similar to the first but it weighted the independent predictors by 1 unit for each 0.3 increase in β-coefficient, thus resulting in a 16-point scoring system (Table 2).

Estimation of risk score

One of the study goals was to develop an easy-to-use prognostic score for stratifying late PSE risk on the basis of a

```
┌─────────────────────────────────────────┐
│ Patients who had an initial ischemic     │
│ stroke hospitalization                   │
│ between the years 2003 and 2014          │
│ N=192,801                                │
└─────────────────────────────────────────┘
```

┌───┐
│ 17,215 Excluded   │
│ 1. Age <45 years (N=9,603)   │
│ 2. Had unknown sex and not citizen of Taiwan (N=531)   │
│ 3. Died within stroke hospitalization (N=92)   │
│ 4. Had a history of epilepsy before initial stroke onset   │
│   (N=3,731)   │
│ 5. Had a history of hospital stay >1 month (N=3,258)   │
└───┘

┌───┐
│ Acute ischemic stroke patients without   │
│ history of epilepsy   │
│ N=175,586   │
└───┘

┌───┐
│ 49,829 Excluded due to a history of TBI, metastatic   │
│ carcinoma, brain tumor, cerebral palsy, anoxic brain injury, │
│ encephalopathy, alchohol abuse, drug abuse, CNS infections   │
│ during the study period   │
└───┘

┌───┐
│ Acute ischemic stroke patients without   │
│ common risk factors associated with   │
│ epilepsy occurrence   │
│ N=125,757   │
└───┘

┌──────────────────────┐   ┌──────────────────────┐
│ Development cohort   │   │ Validation cohort   │
│ Jan 2003–Dec 2010   │   │ Jan 2011–Dec 2014   │
│ N=87,068   │   │ N=38,689   │
└──────────────────────┘   └──────────────────────┘

┌──────────────────────┐   ┌──────────────────────┐
│ 1-year PSE   │   │ 1-year PSE   │
│ N=1,684   │   │ N=725   │
└──────────────────────┘   └──────────────────────┘

Figure 1 Flowchart of sample selection.
Abbreviations: CNS, central nervous system; PSE, poststroke epilepsy; TBI, traumatic brain injury.

Table 2 Independent predictors of 1-year late PSE risk based on backward elimination Cox regression and corresponding risk scores

Independent variables	β-coefficient	(SE)	HR	(95% CI)	Risk scoring system	
					Schneeweiss	Nearest integer
Prolonged hospital staying (>2 weeks)	0.42	(0.06)	1.52	(1.36–1.70)	1	0
Seizure at stroke admission	1.76	(0.06)	5.81	(5.20–6.48)	6	2
Elderly patients (age ≥80 years)	0.15	(0.06)	1.17	(1.04–1.31)	1	0
ICU stay	1.05	(0.06)	2.86	(2.54–3.22)	3	1
Cognitive impairment (dementia)	0.62	(0.11)	1.85	(1.48–2.32)	2	1
Atrial fibrillation	0.55	(0.06)	1.74	(1.55–1.95)	2	1
Respiratory tract infection (pneumonia) on stroke admission	0.27	(0.07)	1.30	(1.14–1.49)	1	0
Harrell's *C*	(0.789)	–	–	–	(0.787)	(0.758)

Abbreviations: ICU, intensive care unit; PSE, poststroke epilepsy.

risk score. Therefore, we classified the patients into low-, medium-, high-, and very-high-risk groups (Table 3). In the development cohort, late PSE incidence (100 person-years) was 0.64 (95% CI: 0.56–0.71) in the low-risk, 2.62 (95% CI: 2.43–2.82) in the medium-risk, 10.3 (95% CI: 9.48–11.3) in the high-risk, and 28.2 (95% CI: 24.0–33.0) in the very-high-risk group per 100 person-years. There was no significant difference in incidence between the development and validation cohorts, except in the low-risk group (0.40/100 person-years, 95% CI: 0.32–0.50). The Kaplan–Meier curve showed that late PSE risk was significantly different in the risk groups based on the PSEiCARe score (Figure 2A and B).

Model performance

The model performance of the Schneeweiss risk scoring system was assessed using Harrell's *C* index, which was 0.762 and 0.792 in the development and validation cohorts, respectively. The optimism-corrected value of Harrell's *C* index was 0.759. Thus, the late PSE prediction model demonstrated fairly favorable internal and external discrimination. The performance of the model was slightly lower when using the nearest integer as the risk score, but it still had moderate discrimination.

Discussion

On the basis of a large population-based health care claims cohort, this study developed the PSEiCARe score for predicting 1-year late PSE risk among patients with ischemic stroke. The PSEiCARe score demonstrated fairly good discrimination and was validated internally through bootstrapping and externally by temporally dividing the data, with consistent findings obtained.

Managing PSE is crucial for stroke care. The guidelines for acute ischemic and hemorrhagic stroke management do not recommend routine administration of prophylactic AEDs,[12,13,24] primarily because few clinical trials have demonstrated a beneficial prophylaxis of AEDs in PSE. Another major point is that most patients with acute stroke do not develop PSE; therefore, adverse effects may outweigh treatment benefits if prophylactic AEDs are administered to all patients. The concept of routinely administering prophylactic AEDs to "very high risk" patients has been proposed.[12,24] Clinicians and researchers could identify high-risk patients who might benefit from such treatment. The PSEiCARe scoring system developed in this study can accurately predict late PSE risk; however, this study was unable to define a cutoff PSEiCARe score for recommending that patients receive AEDs. This problem requires further investigation to weigh

Table 3 Incidence (100 person-years) and adjusted HR of 1-year late PSE risk among the 4 risk groups

Scoring system	Risk group	Score	Development cohort				Validation cohort			
			No of patients	No of PSE patients	Incidence (95% CI)	Adjusted HR (95% CI)	No of patients	No of PSE patients	Incidence (95% CI)	Adjusted HR (95% CI)
Schneeweiss	Low	0	45,781	283	0.64 (0.56–0.71)	1.00 (Ref)	19,031	75	0.40 (0.32–0.50)	1.00 (Ref)
	Medium	1–5	32,949	719	2.62 (2.43–2.82)	4.07 (3.54–4.66)	15,845	313	2.28 (2.04–2.55)	5.63 (4.38–7.24)
	High	6–10	7,242	525	10.3 (9.48–11.3)	15.8 (13.7–18.3)	3,296	258	10.9 (9.62–12.3)	26.5 (20.5–34.3)
	Very high	≥11	1,096	157	28.2 (24.0–33.0)	41.3 (34.0–50.2)	517	79	28.2 (22.3–35.2)	66.4 (48.4–91.2)
	Harrell's C	–	–	–	–	(0.762)	–	–	–	(0.792)
	Optimism correction	–	–	–	–	(0.759)	–	–	–	–
Nearest integer	Low	0	61,807	508	0.86 (0.79–0.94)	1.00 (Ref)	26,135	156	0.62 (0.53–0.73)	1.00 (Ref)
	Medium	1	16,318	455	3.61 (3.29–3.96)	4.12 (3.63–4.68)	8,270	209	3.08 (2.68–3.53)	4.92 (4.00–6.05)
	High	2–3	8,389	651	11.2 (10.4–12.1)	12.68 (11.3–14.2)	3,989	315	10.8 (9.66–12.1)	17.1 (14.1–20.7)
	Very high	≥4	554	70	23.2 (18.1–29.3)	25.49 (19.9–32.7)	295	45	27.4 (20.0–36.6)	42.1 (30.2–58.7)
	Harrell's C	–	–	–	–	(0.755)	–	–	–	(0.789)
	Optimism correction	–	–	–	–	(0.753)	–	–	–	–

Abbreviation: PSE, poststroke epilepsy.

Figure 2 Kaplan–Meier curve for cumulative risk of PSE within 1 year after stroke in the development (**A**) and validation (**B**) cohorts.
Abbreviation: PSE, poststroke epilepsy.

the health and economic benefits of AEDs in reducing PSE occurrence, as well as possible complications stemming from their use.

The independent predictors of late PSE in this study – including prolonged hospital stay, pneumonia, and atrial fibrillation – are preventable and manageable, as the acute stroke management guidelines indicate. Atrial fibrillation is the most common cause of cardioembolic stroke and is associated with a higher chance of cortical ischemic lesion, spontaneous arterial recanalization, and hemorrhagic transformation compared with ischemic strokes of other etiologies.[25–27] Therefore, managing atrial fibrillation is vital because of its preventive effects for both stroke recurrence and PSE.

Acute ischemic stroke management guidelines recommend thrombolytic therapy for eligible patients to reduce their symptom severity and LOS.[13] Airway protection and dysphagia screening are recommended to reduce the likelihood of aspiration pneumonia,[13,28] which may also shorten LOS. Additionally, antibiotics such as cephalosporin decrease the threshold of epileptic seizures,[29] which might be another reason why pneumonia increases the risk of PSE. Therefore, prevention is more important than treatment of pneumonia in stroke patients. Guideline-directed management of patients with stroke improves their neurological outcomes,[30,31] and it may also be a more practical approach to reducing PSE risk than prophylactic AEDs.

The PSEiCARe score is advantageous because the required information is easily obtained from medical records. A previously proposed PSE risk scale based on data from a single hospital revealed that 7 of 10 clinical characteristics were associated with PSE risk.[6] However, the model was not

validated internally or externally because of the single-center study design with relatively small sample size. By contrast, the present study used a larger sample size, resulting in a more precise prediction model. The model developed in this study used information available in medical records and insurance claims data. The model's performance was in accordance with the TRIPOD guidelines. Therefore, physicians and researchers can use the PSEiCARe scoring model in clinical practice to easily identify patients with different PSE risk levels.

A prolonged hospital stay and pneumonia may be interrelated. Models without checking for multicollinearity among variables may lead to erroneous system analysis. Fortunately, this problem can be eliminated by using appropriate statistical methods, such as the stepwise regression used in this study, to select appropriate predictors from the data set. In this study, prolonged hospital stay and respiratory tract infection on stroke admission were both associated with the risk of late PSE. These 2 variables are highly likely to be multicollinear (Spearman rank correlation coefficient =0.297). To better understand how the model would change if 1 of these 2 variables was excluded, we performed a separate analysis, and the results of this analysis showed that excluding a variable did not alter the model (Table S2). The original model had superior performance, and we therefore suggest retaining both variables in the predictive model.

This study has several limitations. First, the model was developed using data from population-based health care claims, and the quality of such data may have influenced our findings. Therefore, we required discharge claims and at least 2 diagnostic claims from outpatient visits within the observational period to confirm the existence of disease; this

approach has been validated and widely used in claims-based observational studies.[32–36] However, potential bias could have occurred due to misclassification. Fortunately, we were more likely to underestimate the effect of risk on late PSE, given that misclassification was nondifferential for patients both with and without PSE. The results should not be changed even if we underestimated the PSE risk. Moreover, because the exact time of seizure onset was not recorded in the claims data, it is possible that a seizure occurring 7 days after a stroke was misclassified as "early seizure" if the patient remained hospitalized for >7 days. Nevertheless, 86% of early seizures were reported to occur within 3 days of a stroke, with 66% occurring within the first 24 hours.[37] Therefore, we believe that this approach should be compatible with the ILAE recommendations.

Second, similar to all prediction models, our results may be biased toward the population in which the model was developed and, consequently, may yield less accurate results in a different population. We assessed the internal validity of the model using a bootstrapping procedure to correct overoptimism bias and also validated the model externally by temporally dividing the data (ie, forming a different population in the same country), which may have mitigated the influence of time-dependent factors. However, external validation of the model with other ethnicities and countries is required to establish its generalizability.

Third, the NHIRD does not include information on lesion location for patients with stroke. Therefore, although cortical lesion is a known PSE risk factor, we were unable to include it in the prediction model. Although recording such detailed information in a population-based database is difficult, the missing information on lesion location in this study may have lowered the model's predictive power. However, to our knowledge, no study has proved that the influence of the cortical stroke factor is stronger than that of other risk factors. Furthermore, including stroke location in the data could have caused problems in a multicenter study, including interrater variation in imaging interpretation and heterogeneity of imaging modality and quality between different hospitals. This may explain why previous studies of PSE were more likely to be single-center studies when stroke location was evaluated.[5,6] Moreover, the generalizability of a model developed in a single-center study is limited. Nevertheless, we recognize that the PSE prediction model developed in this study would be improved if stroke location were included.

Fourth, the observation period was 12 years, which might cause concern that care had changed over time. Nevertheless, no significant difference in stroke care occurred. The guidelines published by the American Heart Association and the American Stroke Association, which are usually referred to by neurologists in Taiwan, were revised in 2003, 2007, and 2013;[13,38,39] most of the changes concerned thrombolytic therapy, which would have had little effect on the PSE risk factors proposed in this study.

Finally, this study only predicted the risk of PSE within 1 year of stroke onset because stroke survivors typically undergo aggressive rehabilitation for at least 6 months to 1 year;[40] therefore, PSE has the most deleterious effect on stroke recovery during this period. However, PSE can also occur >1 year after stroke onset.

Conclusion

This study used population-based health care claims and applied a multivariate prediction model to develop the PSEi-CARe score for predicting 1-year late PSE risk. This score classifies late PSE risk level on the basis of patient disease condition on stroke admission. Physicians might consider this score as an additional tool for improving late PSE prediction. The major late PSE risk factors are the known comorbidities of acute stroke, which are preventable and manageable. Therefore, guideline-based management of acute stroke is crucial for preventing late PSE. External validation is warranted to confirm our findings in different populations.

Acknowledgments

Disclosure

The authors report no conflicts of interest in this work.

References

1. Balami JS, Chen RL, Grunwald IQ, Buchan AM. Neurological complications of acute ischaemic stroke. *Lancet Neurol*. 2011;10(4):357–371.
2. Kwan J. Stroke: predicting the risk of poststroke epilepsy-why and how? *Nat Rev Neurol*. 2010;6(10):532–533.
3. Leidy NK, Elixhauser A, Vickrey B, Means E, Willian MK. Seizure frequency and the health-related quality of life of adults with epilepsy. *Neurology*. 1999;53(1):162–166.
4. Myint PK, Staufenberg EFA, Sabanathan K. Post-stroke seizure and post-stroke epilepsy. *Postgrad Med J*. 2006;82(971):568–572.
5. Lossius MI, Ronning OM, Slapo GD, Mowinckel P, Gjerstad L. Poststroke epilepsy: occurrence and predictors--a long-term prospective controlled study (Akershus Stroke Study). *Epilepsia*. 2005;46(8):1246–1251.
6. Strzelczyk A, Haag A, Raupach H, Herrendorf G, Hamer HM, Rosenow F. Prospective evaluation of a post-stroke epilepsy risk scale. *J Neurol*. 2010;257(8):1322–1326.

7. Camilo O, Goldstein LB. Seizures and epilepsy after ischemic stroke. *Stroke*. 2004;35(7):1769–1775.

8. Ferlazzo E, Gasparini S, Beghi E, et al; Epilepsy Study Group of the Italian Neurological Society. Epilepsy in cerebrovascular diseases: review of experimental and clinical data with meta-analysis of risk factors. *Epilepsia*. 2016;57(8):1205–1214.

9. Couillard P, Almekhlafi MA, Irvine A, et al. Subacute seizure incidence in thrombolysis-treated ischemic stroke patients. *Neurocrit Care*. 2012;16(2):241–245.

10. Chiang IH, Chang WN, Lin WC, et al. Risk factors for seizures after first-time ischemic stroke by magnetic resonance imaging. *Acta Neurol Taiwan*. 2010;19(1):26–32.

11. Rhoney DH, Tipps LB, Murry KR, Basham MC, Michael DB, Coplin WM. Anticonvulsant prophylaxis and timing of seizures after aneurysmal subarachnoid hemorrhage. *Neurology*. 2000;55(2):258–265.

12. Connolly ES Jr, Rabinstein AA, Carhuapoma JR, et al; American Heart Association Stroke Council; Council on Cardiovascular Radiology and Intervention; Council on Cardiovascular Nursing; Council on Cardiovascular Surgery and Anesthesia; Council on Clinical Cardiology. Guidelines for the management of aneurysmal subarachnoid hemorrhage: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2012;43(6):1711–1737.

13. Jauch EC, Saver JL, Adams HP Jr, et al; American Heart Association Stroke Council; Council on Cardiovascular Nursing; Council on Peripheral Vascular Disease; Council on Clinical Cardiology. Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2013;44(3):870–947.

14. Silverman IE, Restrepo L, Mathews GC. Poststroke seizures. *Arch Neurol*. 2002;59(2):195–201.

15. Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338:b606.

16. Steyerberg EW, Harrell FE Jr, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDF. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. 2001;54(8):774–781.

17. Maaijwee NA, Rutten-Jacobs LC, Schaapsmeerders P, van Dijk EJ, de Leeuw FE. Ischaemic stroke in young adults: risk factors and long-term consequences. *Nat Rev Neurol*. 2014;10(6):315–325.

18. Thurman DJ, Beghi E, Begley CE, et al; ILAE Commission on Epidemiology. Standards for epidemiologic studies and surveillance of epilepsy. *Epilepsia*. 2011;52(suppl 7):2–26.

19. Fischer U, Arnold M, Nedeltchev K, et al. Impact of comorbidity on ischemic stroke outcome. *Acta Neurol Scand*. 2006;113(2):108–113.

20. Mehta HB, Mehta V, Girman CJ, Adhikari D, Johnson ML. Regression coefficient-based scoring system should be used to assign weights to the risk index. *J Clin Epidemiol*. 2016;79:22–28.

21. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA*. 1982;247(18):2543–2546.

22. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361–387.

23. Moons KGM, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): explanation and elaboration the TRIPOD statement: explanation and elaboration. *Ann Intern Med*. 2015;162(1):W1–W73.

24. Morgenstern LB, Hemphill JC 3rd, Anderson C, et al. Guidelines for the management of spontaneous intracerebral hemorrhage: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2010;41(9):2108–2129.

25. Paciaroni M, Agnelli G, Corea F, et al. Early hemorrhagic transformation of brain infarction: rate, predictive factors, and influence on clinical outcome: results of a prospective multicenter study. *Stroke*. 2008;39(8):2249–2256.

26. Molina CA, Montaner J, Abilleira S, et al. Timing of spontaneous recanalization and risk of hemorrhagic transformation in acute cardioembolic stroke. *Stroke*. 2001;32(5):1079–1084.

27. Petersen P, Madsen EB, Brun B, Pedersen F, Gyldensted C, Boysen G. Silent cerebral infarction in chronic atrial fibrillation. *Stroke*. 1987;18(6):1098–1100.

28. Donovan NJ, Daniels SK, Edmiaston J, et al; American Heart Association Council on Cardiovascular Nursing and Stroke Council. Dysphagia screening: state of the art: invitational conference proceeding from the State-of-the-Art Nursing Symposium, International Stroke Conference 2012. *Stroke*. 2013;44(4):e24–e31.

29. Grill MF, Maganti R. Cephalosporin-induced neurotoxicity: clinical manifestations, potential pathogenic mechanisms, and the role of electroencephalographic monitoring. *Ann Pharmacother*. 2008;42(12):1843–1850.

30. Hsieh FI, Lien LM, Chen ST, et al; Taiwan Stroke Registry Investigators. Get with the guidelines-stroke performance indicators: surveillance of stroke care in the Taiwan stroke registry: get with the guidelines-stroke in Taiwan. *Circulation*. 2010;122(11):1116–1123.

31. Schwamm LH, Fonarow GC, Reeves MJ, et al. Get with the guidelines-stroke is associated with sustained improvement in care for patients hospitalized with acute stroke or transient ischemic attack. *Circulation*. 2009;119(1):107–115.

32. Cheng CL, Kao YH, Lin SJ, Lee CH, Lai ML. Validation of the National Health Insurance Research Database with ischemic stroke cases in Taiwan. *Pharmacoepidemiol Drug Saf*. 2011;20(3):236–242.

33. Chi NF, Chien LN, Ku HL, Hu CJ, Chiou HY. Alzheimer disease and risk of stroke: a population-based cohort study. *Neurology*. 2013;80(8):705–711.

34. Abbas S, Ihle P, Koster I, Schubert I. Estimation of disease incidence in claims data dependent on the length of follow-up: a methodological approach. *Health Serv Res*. 2012;47(2):746–755.

35. Helmers SL, Thurman DJ, Durgin TL, Pai AK, Faught E. Descriptive epidemiology of epilepsy in the U.S. population: a different approach. *Epilepsia*. 2015;56(6):942–948.

36. Margolis JM, Chu BC, Wang ZJ, Copher R, Cavazos JE. Effectiveness of antiepileptic drug combination therapy for partial-onset seizures based on mechanisms of action. *JAMA Neurol*. 2014;71(8):985–993.

37. Reith J, Jorgensen HS, Nakayama H, Raaschou HO, Olsen TS. Seizures in acute stroke: predictors and prognostic significance. The Copenhagen Stroke Study. *Stroke*. 1997;28(8):1585–1589.

38. Adams HP Jr, Adams RJ, Brott T, et al; Stroke Council of the American Stroke Association. Guidelines for the early management of patients with ischemic stroke: a scientific statement from the stroke council of the American Stroke Association. *Stroke*. 2003;34(4):1056–1083.

39. Adams HP Jr, del Zoppo G, Alberts MJ, et al; American Heart Association/American Stroke Association Stroke Council; American Heart Association/American Stroke Association Clinical Cardiology Council; American Heart Association/American Stroke Association Cardiovascular Radiology and Intervention Council; Atherosclerotic Peripheral Vascular Disease Working Group; Quality of Care Outcomes in Research Interdisciplinary Working Group. Guidelines for the early management of adults with ischemic stroke: a guideline from the American Heart Association/American Stroke Association Stroke Council, Clinical Cardiology Council, Cardiovascular Radiology and Intervention Council, and the Atherosclerotic Peripheral Vascular Disease and Quality of Care Outcomes in Research Interdisciplinary Working Groups: the American Academy of Neurology affirms the value of this guideline as an educational tool for neurologists. *Stroke*. 2007;38(5):1655–1711.

40. Miller EL, Murray L, Richards L, et al; American Heart Association Council on Cardiovascular Nursing and the Stroke Council. Comprehensive overview of nursing and interdisciplinary rehabilitation care of the stroke patient: a scientific statement from the American Heart Association. *Stroke*. 2010;41(10):2402–2448.

Supplementary materials

Table S1 Disease diagnostic coding

Disease	ICD-9-CM codes
Stroke subtype	
Ischemic	433–434, 436
Subarachnoid hemorrhage	430
Intracerebral hemorrhage	431
Epilepsy	345, 780.3, 780.39
Pneumonia admission within 14 days after stroke	480–486, 507
Comorbidities	
Hypertension	401–402
Diabetes	250
Hyperlipidemia	272
Coronary artery disease	413–414, 429.2
Heart failure	428
Atrial fibrillation	427.3
Peripheral artery disease	440.2–440.4, 443.9
Malignant neoplasm	140–239
Renal disease	403.01, 403.11, 403.91, 404.02–404.03, 404.12–404.13, 404.92, 404.93, 582, 583–583.2, 583.4, 583.6–583.7, 585–586, 588, V42, V45.1, V56
Moderate or severe liver disease	456–456.2, 572.2–572.4, 572.8
Rheumatoid arthritis or collagen vascular disease	446, 701, 710–710.4, 710.8–710.9, 711.2, 714, 719.3, 720, 725, 728.5, 728.89, 729.3
Dementia	290, 294.1, 331.0, 331.2
Parkinson disease	332
Exclusive disease conditions	
Metastatic carcinoma	196–199
Brain tumor	191–192, 194.3–194.4, 200.5, 225–225.2, 225.8–225.9, 227.3–227.4, 237–237.1, 237.5–237.7, 237.9, 239.6
Traumatic brain and head injuries	850–854, 959.01
Cerebral palsy	343
Anoxic brain injury	348.1, 768.5–768.9
Encephalopathy	348.3
Alcohol abuse	265.2, 291.1–291.3, 291.5, 291.8, 291.9, 303, 303.9, 305, 357.5, 425.5, 535.3, 571–571.3, 980, V11.3
Drug abuse	292, 304, 305.2–305.9, V65.42
Central nervous system infections	320–326

Abbreviation: ICD-9-CM, International Classification of Diseases, Ninth Revision, Clinical Modification.

Table S2 Comparison of β-coefficients among different models

Independent variables	Different model selection								
	Model with all predictors			Model without prolonged hospital stay			Model without respiratory tract infection on stroke admission		
	β-coefficient	Risk scoring system		β-coefficient	Risk scoring system		β-coefficient	Risk scoring system	
		β/ Schneeweiss	Nearest integer		β/ Schneeweiss	Nearest integer		β/ Schneeweiss	Nearest integer
Prolonged hospital stay (>2 weeks)	0.42	1	0				0.46	2	0
Seizure at stroke admission	1.76	6	2	1.81	6	2	1.77	6	2
Elderly patients (age ≥80 years)	0.15	1	0	0.17	1	0	0.18	1	0
ICU stay, yes	1.05	3	1	1.20	4	1	1.10	4	1
Cognitive impairment (dementia)	0.62	2	1	0.62	2	1	0.62	2	1
Atrial fibrillation	0.55	2	1	0.58	2	1	0.56	2	1
Respiratory tract infection (pneumonia) on stroke admission	0.27	1	0	0.36	1	0			
Harrell's C	(0.789)	(0.787)	(0.758)	(0.776)	(0.773)		(0.786)	(0.784)	
Total score		16	5		16	5		17	5

Abbreviation: ICU, intensive care unit.